Changing Relations. A GMM approach to W

Majed Atwi[1], Jan Hauke[2],Tomasz Kossowski[2] and Jesús Mur[1]

[1]University of Zaragoza (Spain)
[2]University of Poznan (Poland)

Abstract

Spatial models deals primarily with unobserved phenomena such as spillovers, transboundary competition, knowledge flows, etc. However, rarely does the user know about how these events operate in practice. This is a problem because these effects are required to build the model. Traditionally the gap has been solved by providing externally this information, in the form of a weighting matrix which must be specified beforehand. It is obvious that further analysis is contingent on that decision. Our purpose is to offer some help to the user in the crucial decision of building a weighting matrix for his/her spatial equation using a Generalized Method of Moments approach.

1 Introduction

Spatial models deal frequently with unobserved phenomena such as spillovers effects, transboundary competition or cooperation relations, knowledge and other intangible flows. The user rarely knows about how these events operate in practice, although very often they are key elements in the specification but they are required to build the model. Traditionally the gap has been solved by providing externally the required information, in the form of a weighting matrix that reflects some priors from the user.

Although reasonable, there has been little consensus on the suitability of this solution, and this raised the **W** issue. The debate became stronger after the work of Ord (1975) because of its strong emphasis on the task of modelling spatial relationships. It is evident that, in order to model, we need good proxies to describe the way agents interact accross space. Later on, Anselin (1988, 2002) put the **W** issue in the very centre of the debate about specification of spatial models, which requires something more robust than just user beliefs. The pressure for more consistent spatial specifications favoured the appearance of an increasingly specialized literature on **W** based on more objective fundaments.

The purpose of the **W** issue is clear: we have to '*to define for any set of points or area objects the spatial relationships that exist between them*' as stated by Haining (2003, p.74). The problem is how should that be done.

Roughly, we may distinguish two approaches to this question: (i) specifying **W** exogenously; (ii) estimating **W** from data. The exogenous approach is by far the most common and includes, for example, use of a binary contiguity criterion, k-nearest neighbours, kernel functions based on distance, etc. The second approach uses the topology of the space and the nature of the data, and takes many forms. We find ad-hoc procedures in which a predefined objective guides the search such as the maximization of Moran's I in Kooijman (1976) or the local statistical model of Getis and Aldstadt (2004). Benjanuvatra and Burridge (2015) develop a quasi maximum-likelihood, QML, algorithm to estimate the weights in **W** assuming partial knowledge about the form of the weights. More flexible approaches are possible if we have panel information such as in Battacharjee and Jensen-Butler (2013) or Beenstock and Felsenstein (2012). Endogeneity of the weight matrix is another topic introduced recently by Kelejian and Piras (2014) and Qu and Lee (2015). This is a promising avenue for the debate about **W**, which connects with the concept of *coevolution* put forward by Snijders et *al* (2007) whose basic assumption is difficult to object: in the long run, the connectivity of the network must evolve with the rest of the system and, in particular, with the endogenous variable of the model. The recent literature on spatial econometrics is developping according to the second approach, but most part of the applied research still relies on the exogenous approach.

Another question of concern in this debate come from the criticisms of Gibbons and Overman (2012), when they say that most spatial models are not identified because the network where the agents interact are not identified. The only way to attain identification is through the specification of a exogenous $\mathbf{W}$ weighting matrix, which is the cause of frequent (weak) identification problems in the models; this flaw extends to the instruments, moment conditions, etc. There is little to say in relation to this point that we also share; in fact, in spite of specifying $\mathbf{W}$ exogenously, there is an overlapping between the autocorrelation coefficients (i.e., the typical ρ parameter) and the pattern in which the interaction happens (the matrix $\mathbf{W}$). The two elements are only jointly identified (Hays et *al.,* 2010) and can not be separated. In fact, this mixture is in the origin of the discussion about what is the most adequate version of the weighting matrix (raw data, row-standardized, eigenvalue-standardized, etc).

Our purpose is to offer some help to the user in the crucial decision of building a weighting matrix for his/her spatial equation. We show that in cases where the user has a panel database with a long time series and small cross-sections and the assumption of symmetry of the weights of W is reasonable, this matrix can be extracted from the covariance matrix of the model (either that of the endogenous variable or from the residuals). Section 2 discusses the unequivocal relation that exists in a typical spatial econometric model between the covariance matrix and the weighting matrix, which allows us, under specific circumstances, to obtain the second matrix from the former. Section 3 formalizes that discussion in a GMM framework. Section 4, shows the results of a small Monte Carlo experiment solved in order to check the results put forward in relation to the GMM approach. Section 5 concludes.

2 The Covariance and the Weighting matrix

Let us assume that we have a panel dataset made of n spatial units observed during T periods. In that panel we have information of variable y and k covariates, x. The sequence is complete; that is, there are no missing data. Moreover, we are pretty sure that there is spatial structure in the sample because, for example, the CD test of Pesaran (2004) has rejected the null hypothesis of no correlation among the cross-sectional units. The hypothesis of common factors (strong spatial correlation) has been discarded and we are thinking in terms of mutual influences between the spatial units (weak spatial correlation). As said, the literature on spatial econometrics (Harris et *al.*, 2011) advocates for the use of the so-called weighting matrix to capture the network of cross-sectional dependences. This matrix is required to advance wiht the data in order to specify and estimate a model for y. However, it is unusual to have enough, uncontroversial information to build the matrix (i.e., Corrado and Fingleton, 2012). Our suggestion is to exploit the direct link between the covariance and the weighting matrix to approach the last one using information in the covariances. Under certain circumstances, this can be done quite directly.

Let us begin with the case of an spatial autorregresive (SAR) process, stable across time:

$$y_t = \rho \mathbf{S} y_t + u_t;\ u_t \sim i.i.d.\,(0; \mathbf{\Sigma})\,;\ t = 1, 2, ..., T \tag{1}$$

y_t is a $(n \times 1)$ vector observed in period t whereas $\mathbf{S}$ is a $(n \times n)$ matrix of unobserved weights and $\mathbf{\Sigma}$ a diagonal matrix reflecting the heteroskedastic nature of the error terms, $\mathbf{\Sigma} = diag\,(\sigma_1^2; \sigma_2^2; ...; \sigma_n^2)$. The covariance matrix of vector y is:

$$\mathbf{\Gamma}_y = (\mathbf{I} - \rho\mathbf{S})^{'-1}\, \mathbf{\Sigma}\, (\mathbf{I} - \rho\mathbf{S})^{-1} \tag{2}$$

The sampling covariance, $\left(\widehat{\mathbf{\Gamma}}_y = \Sigma_{t=1}^{T} y_t y_t^{'}/T\right)$, under these circumstances, is a consistent estimate of $\mathbf{\Gamma}_y$, so $p\lim \widehat{\mathbf{\Gamma}} = \mathbf{\Gamma}_y$. The sampling covariance contains $\frac{n(n+1)}{2}$ different estimates whereas the number of unknowns in (2) is potentially larger: n variances in $\mathbf{\Sigma}$, $n(n-1)$ terms in $\mathbf{S}$ (as usual, we assume, that the diagonal terms are zero, $s_{ii} = 0; i = 1; ...; n)$. As said, if the weighting matrix is unknown, the spatial dependence parameter ρ is not identified; to simplfy, in the following we assume that this parameter is subsumed into the weights so that $\rho\mathbf{S} = \mathbf{W}$. Without loss of generality, in the following we assume that the dependence parameter is embodied in the weighting matrix. Another common assumption in applied work is that the weights in $\mathbf{W}$ are symmetric (this assumption is not without criticism; indeed, symmetry simplifies the treatment of the data at the cost of assuming an often no realistic assumption). In sum, the number of unknowns in the right hand side of (2) is exactly $\frac{n(n+1)}{2}$: there are n variances and $\frac{n(n-1)}{2}$ different weights in $\mathbf{W}$. This means that the spatial weights and the variances can be recovered from the sampling covariances.

As an example, let us assume a very simple case in which $n = 4$. The sampling covariance is a (4×4) matrix which allows to write the inverse of (2) as:

$$\mathbf{\Gamma}_y^{-1} = (\mathbf{I} - \mathbf{W})\, \mathbf{\Sigma}^{-1}\, (\mathbf{I} - \mathbf{W}) \tag{3}$$

We obtain the following system of nonlinear equations:

$$\left.\begin{array}{r}
\underline{\gamma}_{11} = \frac{1}{\sigma_1^2} + \frac{\omega_{12}^2}{\sigma_2^2} + \frac{\omega_{13}^2}{\sigma_3^2} + \frac{\omega_{14}^2}{\sigma_4^2} \\
\underline{\gamma}_{22} = \frac{1}{\sigma_2^2} + \frac{\omega_{21}^2}{\sigma_1^2} + \frac{\omega_{23}^2}{\sigma_3^2} + \frac{\omega_{24}^2}{\sigma_4^2} \\
\underline{\gamma}_{33} = \frac{1}{\sigma_3^2} + \frac{\omega_{31}^2}{\sigma_1^2} + \frac{\omega_{31}^2}{\sigma_2^2} + \frac{\omega_{34}^2}{\sigma_4^2} \\
\underline{\gamma}_{44} = \frac{1}{\sigma_4^2} + \frac{\omega_{41}^2}{\sigma_1^2} + \frac{\omega_{41}^2}{\sigma_2^2} + \frac{\omega_{43}^2}{\sigma_3^2} \\
\underline{\gamma}_{12} = -\frac{\omega_{12}}{\sigma_1^2} - \frac{\omega_{12}}{\sigma_2^2} + \frac{\omega_{13}\omega_{23}}{\sigma_3^2} + \frac{\omega_{14}\omega_{24}}{\sigma_4^2} \\
\underline{\gamma}_{13} = -\frac{\omega_{13}}{\sigma_1^2} - \frac{\omega_{13}}{\sigma_3^2} + \frac{\omega_{12}\omega_{32}}{\sigma_3^2} + \frac{\omega_{14}\omega_{34}}{\sigma_4^2} \\
\underline{\gamma}_{14} = -\frac{\omega_{14}}{\sigma_1^2} - \frac{\omega_{14}}{\sigma_4^2} + \frac{\omega_{12}\omega_{24}}{\sigma_2^2} + \frac{\omega_{13}\omega_{43}}{\sigma_4^2} \\
\underline{\gamma}_{23} = -\frac{\omega_{23}}{\sigma_2^2} - \frac{\omega_{23}}{\sigma_3^2} + \frac{\omega_{21}\omega_{31}}{\sigma_1^2} + \frac{\omega_{24}\omega_{34}}{\sigma_4^2} \\
\underline{\gamma}_{24} = -\frac{\omega_{24}}{\sigma_2^2} - \frac{\omega_{24}}{\sigma_4^2} + \frac{\omega_{12}\omega_{41}}{\sigma_1^2} + \frac{\omega_{23}\omega_{43}}{\sigma_4^2} \\
\underline{\gamma}_{34} = -\frac{\omega_{34}}{\sigma_3^2} - \frac{\omega_{34}}{\sigma_4^2} + \frac{\omega_{31}\omega_{41}}{\sigma_2^2} + \frac{\omega_{32}\omega_{42}}{\sigma_4^2}
\end{array}\right\} \tag{4}$$

where $\underline{\gamma}_{ij}$ is the $(i;j)$ element of $\mathbf{\Gamma}_y^{-1}$, σ_j^2 is the jth variance in $\mathbf{\Sigma}$ and ω_{ij} the corresponding (symmetric) weight in $\mathbf{W}$. This is a system of 10 equations and 10 unknows, with a unique solution (as said, the system is compatible and determined). Using the sampling covariances, the corresponding estimates of $\mathbf{\Sigma}$ and $\mathbf{W}$ are consistent as long as $\widehat{\mathbf{\Gamma}}_y$ remains as a consistent estimator of $\mathbf{\Gamma}_y$. For a more general case of n spatial units, we obtain:

$$\left.\begin{array}{l}
\underline{\gamma}_{ii} = \frac{1}{\sigma_i^2} + \sum_{i\neq j}^n \frac{\omega_{ij}^2}{\sigma_j^2};\ i = 1,2,...,n \\
\underline{\gamma}_{ij} = -\omega_{ij}\left(\frac{1}{\sigma_i^2} + \frac{1}{\sigma_j^2}\right) + \sum_{\substack{l\neq i \\ l\neq j}}^n \frac{\omega_{il}\omega_{jl}}{\sigma_l^2}; i < j
\end{array}\right\} \tag{5}$$

Let us assume now that the dependence appears in the errors of a (linear) model so that:

$$y_t = x_t\beta + u_t;\ u_t = \rho\mathbf{S}u_t + \varepsilon_t; \varepsilon_t \sim i.i.d.\left(0;\mathbf{\Sigma}\right);\ t = 1,2,...,T \tag{6}$$

x_t is a $(n \times k)$ matrix observed in period t and β the corresponding vector of coefficients; ρ remains unidentified so $\rho\mathbf{S} = \mathbf{W}.$ The most favorable situation for our purposes is that the regressors are time stationary (constant mean and variance) and have no spatial structure. Under these circumstances, their impact is limited to the variances in the main diagonal of $\mathbf{\Gamma}_y$ (in other words, the set of covariances are unaltered):

$$\begin{aligned}
\mathbf{\Gamma}_y &= V\left(y_t\right) = V\left(x_t\beta + u_t\right) = V\left(\Sigma_{j=1}^k x_{jt}\beta_j + u_t\right) = \\
&= \Sigma_{j=1}^k \beta_j^2 V\left(x_{jt}\right) + V\left(u_t\right) = \left(\Sigma_{j=1}^k \beta_j^2\sigma_{xj}^2\right) I_n + \mathbf{\Gamma}_u;\ t = 1,...,T
\end{aligned} \tag{7}$$

where σ^2_{xj} is the variance of the j-th regressor and $\mathbf{\Gamma}_u = (\mathbf{I}-\mathbf{W})^{-1}\mathbf{\Sigma}(\mathbf{I}-\mathbf{W})^{-1}$. However, the impact in the variances extends throughout the system of (4), which means that the sample analog of (4) will produce biased estimates of $\mathbf{\Sigma}$ and $\mathbf{W}$. It is clear that the list of exogenous regressors in the right hand side of (6) requires of a preliminay LS estimation to account for their effect. Then, the sampling covariance matrix of (3) should be that of the LS residuals. This preliminary step is required to attain consistency because now there are $\frac{n(n+1)}{2}+2k$ unknowns in (6) and we need of $2k$ additional restrictions: k come from the LS estimation whereas the other k are the sampling variances of the k covariates

The situation becomes intractable if the x covariates are spatially dependent; then:

$$\mathbf{\Gamma}_y = \left(\Sigma^k_{j=1}\beta^2_j\mathbf{\Gamma}_{x_j}\right)+\mathbf{\Gamma}_u;\ t=1,...,T \tag{8}$$

In this case, the covariance matrix of y is a mixture of the covariance matrices of the regressors and the covariance matrix of the error terms.

The case of an Spatial Lag Model (SLM) does not facilitate the discussion, although is solvable under certain circumstances. Assuming that:

$$y_t = \rho\mathbf{S}y_t + x_t\beta + u_t;\ u_t \sim i.i.d.\left(0;\mathbf{\Sigma}\right);\ t=1,2,...,T \tag{9}$$

The covariance matrix of vector y is:

$$\mathbf{\Gamma}_y = (\mathbf{I}-\rho\mathbf{S})^{-1}\left[\left(\Sigma^k_{j=1}\beta^2_j\sigma^2_{xj}\right)I_n+\mathbf{\Sigma}\right](\mathbf{I}-\rho\mathbf{S})^{-1} \tag{10}$$

In fact, the situation is similar to that of (3): the covariance matrix of y, $\mathbf{\Gamma}_y$, is obtained pre and postmultiplying a diagonal matrix, $\mathbf{\Gamma}_{x,u} = \left[\left(\Sigma^k_{j=1}\beta^2_j\sigma^2_{xj}\right)I_n+\mathbf{\Sigma}\right]$, by $(\mathbf{I}-\rho\mathbf{S})^{-1}$. Once again, ρ is not identified so let us call so $\rho\mathbf{S}=\mathbf{W}$. There are $\frac{n(n-1)}{2}$ unknowns in $\mathbf{W}$ and n composite variances in the diagonal of $\mathbf{\Gamma}_{x,u}$. Moreover, we have $\frac{n(n+1)}{2}$ sampling covariances in $\widehat{\mathbf{\Gamma}}_y$, so the system is determined. We will obtain n *composite* variances, such as $\left(\Sigma^k_{j=1}\beta^2_j\sigma^2_{xj}\right)+\sigma^2_i; i=1,2,...,n$ and the $\frac{n(n-1)}{2}$ weights in $\mathbf{W}$. However, it will not be possible to disentangle the variance of the error terms, $\sigma^2_i; i=1,2,...,n$, form the linear combination of the variance of the regressors, $\left(\Sigma^k_{j=1}\beta^2_j\sigma^2_{xj}\right)$. If the regressors are spatially correlated, such as $V(x_j)=\mathbf{\Gamma}_{x_j}$, the problem is not solvable:

$$\mathbf{\Gamma}_y = (\mathbf{I}-\mathbf{W})^{-1}\left[\Sigma^k_{j=1}\beta^2_j\mathbf{\Gamma}_{x_j}+\mathbf{\Sigma}\right](\mathbf{I}-\mathbf{W})^{-1} \tag{11}$$

Finally, let us return to the case of expression (1). In some cases it is possible to assume that the variance of the error terms is the same for the n spatial units so that:

$$y_t = \rho \mathbf{S} y_t + u_t;\ u_t \sim i.i.d.\left(0; \sigma_u^2 I_n\right);\ t = 1, 2, ..., T \tag{12}$$

Proceeding in the same way, the unknowns are now $\frac{n(n-1)}{2}$ symmetric weights in $\mathbf{W} = \rho \mathbf{S}$ plus a common variance, σ_u^2, whereas the number of different terms in the sampling covariance matrix continues to be $\frac{n(n+1)}{2}$. Now the system is overidentified. In fact there are $n-1$ overidentifyng restriccions. Let us write the sample analog of equation (3) for this case

$$\widehat{\mathbf{\Gamma}}_y^{-1} = \widehat{\sigma}_u^{-2}\left(\mathbf{I} - \widehat{\mathbf{W}}\right)\left(\mathbf{I} - \widehat{\mathbf{W}}\right) \tag{13}$$

$\widehat{\sigma}_u^2$ is, as usual, a scale factor for the observed covariances. Rearranging the last expression, we obtain another nonlinear system on the sampling covariances:

$$\left.\begin{array}{r}
\underline{\widehat{\gamma}}_{11} = \frac{1}{\widehat{\sigma}_u^2}(1 + \widehat{\omega}_{12}^2 + \widehat{\omega}_{13}^2 + \widehat{\omega}_{14}^2) \\
\underline{\widehat{\gamma}}_{22} = \frac{1}{\widehat{\sigma}_u^2}(1 + \widehat{\omega}_{21}^2 + \widehat{\omega}_{23}^2 + \widehat{\omega}_{24}^2) \\
\underline{\widehat{\gamma}}_{33} = \frac{1}{\widehat{\sigma}_u^2}(1 + \widehat{\omega}_{31}^2 + \widehat{\omega}_{32}^2 + \widehat{\omega}_{34}^2) \\
\underline{\widehat{\gamma}}_{44} = \frac{1}{\widehat{\sigma}_u^2}(1 + \widehat{\omega}_{41}^2 + \widehat{\omega}_{42}^2 + \widehat{\omega}_{43}^2) \\
\underline{\widehat{\gamma}}_{12} = \frac{1}{\widehat{\sigma}_u^2}(-2\widehat{\omega}_{12} + \widehat{\omega}_{13}\widehat{\omega}_{23} + \widehat{\omega}_{14}\widehat{\omega}_{24}) \\
\underline{\widehat{\gamma}}_{13} = \frac{1}{\widehat{\sigma}_u^2}(-2\widehat{\omega}_{13} + \widehat{\omega}_{12}\widehat{\omega}_{31} + \widehat{\omega}_{14}\widehat{\omega}_{34}) \\
\underline{\widehat{\gamma}}_{14} = \frac{1}{\widehat{\sigma}_u^2}(-2\widehat{\omega}_{14} + \widehat{\omega}_{12}\widehat{\omega}_{42} + \widehat{\omega}_{13}\widehat{\omega}_{43}) \\
\underline{\widehat{\gamma}}_{23} = \frac{1}{\widehat{\sigma}_u^2}(-2\widehat{\omega}_{23} + \widehat{\omega}_{21}\widehat{\omega}_{31} + \widehat{\omega}_{24}\widehat{\omega}_{34}) \\
\underline{\widehat{\gamma}}_{24} = \frac{1}{\widehat{\sigma}_u^2}(-2\widehat{\omega}_{24} + \widehat{\omega}_{21}\widehat{\omega}_{41} + \widehat{\omega}_{23}\widehat{\omega}_{43}) \\
\underline{\widehat{\gamma}}_{34} = \frac{1}{\widehat{\sigma}_u^2}(-2\widehat{\omega}_{34} + \widehat{\omega}_{31}\widehat{\omega}_{1} + \widehat{\omega}_{32}\widehat{\omega}_{42})
\end{array}\right\} \tag{14}$$

We have 10 equation but only 7 unknowns which means that the system is compatible and overidentified. The corresponding estimates from the nonlinear system of (14) will be consistent under the assumption that the sampling covariance matrix, $\widehat{\mathbf{\Gamma}}_y$, is also consistent. The extension to the case of n spatial units, is inmediate:

$$\left.\begin{array}{r}
\underline{\gamma}_{ii} = \frac{1}{\sigma_u^2}(1 + \sum_{i \neq j}^{n} \omega_{ij}^2);\ i = 1, 2, ..., n \\
\underline{\gamma}_{ij} = \frac{1}{\sigma_u^2}(-2\omega_{ij} + \sum_{\substack{l \neq i \\ l \neq j}}^{n} \omega_{il}\omega_{jl}); i < j
\end{array}\right\} \tag{15}$$

where we have $\frac{n(n+1)}{2}$ equations but only $\frac{n(n-1)}{2} + 1$ unknowns.

3 A GMM approach.

Let us assume now a SMA process like the following (results are similar for the previous SAR cases):

$$\begin{array}{c} y_t = \varepsilon_t + \rho \mathbf{S} \varepsilon_t = \varepsilon_t + \mathbf{W} \varepsilon_t;\ \varepsilon_t \sim N\left(0, \mathbf{\Sigma}\right) \\ t = 1, 2, ..., T \end{array} \tag{16}$$

As usual, the weights in the $\mathbf{S}$ symmetric matrix are unknown, which means that ρ is not identified, and we assume a diagonal matrix for the covariances of the error terms, $\mathbf{\Sigma}$, is diagonal. The covariance matrix of vector y_t, $\forall\ t$, is an $(n \times n)$ matrix such as:

$$\Gamma_y = \begin{pmatrix} \sigma_1^2 + \Sigma_{j\neq 1}^n \sigma_j^2 \omega_{ij}^2 & \begin{array}{c}\omega_{12}\left(\sigma_1^2+\sigma_2^2\right)+ \\ \Sigma_{j\neq 1;2}^n \sigma_j^2 \omega_{1j}\omega_{2j}\end{array} & \begin{array}{c}.... \\ ....\end{array} & \begin{array}{c}\omega_{1n}\left(\sigma_1^2+\sigma_n^2\right)+ \\ +\Sigma_{j\neq 1;2}^n \sigma_j^2 \omega_{1j}\omega_{nj}\end{array} \\ \begin{array}{c}\omega_{12}\left(\sigma_1^2+\sigma_2^2\right)+ \\ \Sigma_{j\neq 1;2}^n \sigma_j^2 \omega_{1j}\omega_{2j}\end{array} & \sigma_1^2 + \Sigma_{j\neq 2}^n \sigma_j^2 \omega_{ij}^2 & \begin{array}{c}.... \\ ....\end{array} & \begin{array}{c}\omega_{2n}\left(\sigma_2^2+\sigma_n^2\right)+ \\ +\Sigma_{j\neq 1n2}^n \sigma_j^2 \omega_{2j}\omega_{nj}\end{array} \\ .... & .... & .... & .... \\ \begin{array}{c}\omega_{1n}\left(\sigma_1^2+\sigma_n^2\right)+ \\ \Sigma_{j\neq 1;2}^n \sigma_j^2 \omega_{1j}\omega_{nj}\end{array} & \begin{array}{c}\omega_{2n}\left(\sigma_2^2+\sigma_n^2\right)+ \\ \Sigma_{j\neq 1n2}^n \sigma_j^2 \omega_{2j}\omega_{nj}\end{array} & \begin{array}{c}.... \\ ....\end{array} & \sigma_n^2 + \Sigma_{j=1}^{n-1} \sigma_j^2 \omega_{ij}^2 \end{pmatrix} \tag{17}$$

θ is the $\frac{n(n+1)}{2} \times 1$ vector of unknown parameters for this problem (n variances plus $\frac{n(n-1)}{2}$ weights) such that $\theta = \left(\sigma_1^2, ..., \sigma_n^2, \omega_{12}, ..., \omega_{1n}, \omega_{23} ..., \omega_{n-1,n}\right)'$; θ_0 is the vector of parameters that intervenes in the Data Generating Process, DGP. The same as before, $\widehat{\Gamma}_y$ denotes the sampling covariance matrix (centered if necessary):

$$\widehat{\Gamma}_y = \begin{pmatrix} \Sigma_{t=1}^T y_{1t}^2/T & \Sigma_{t=1}^T y_{1t}y_{2t}/T & .... & \Sigma_{t=1}^T y_{1t}y_{nt}/T \\ \Sigma_{t=1}^T y_{1t}y_{2t}/T & \Sigma_{t=1}^T y_{2t}^2/T & .... & \Sigma_{t=1}^T y_{2t}y_{nt}/T \\ .... & .... & .... & .... \\ \Sigma_{t=1}^T y_{1t}y_{nt}/T & \Sigma_{t=1}^T y_{2t}y_{nt}/T & .... & \Sigma_{t=1}^T y_{nt}^2/T \end{pmatrix} = \begin{pmatrix} \widehat{\gamma}_{11} & \widehat{\gamma}_{12} & .... & \widehat{\gamma}_{1n} \\ \widehat{\gamma}_{12} & \widehat{\gamma}_{22} & .... & \widehat{\gamma}_{2n} \\ .... & .... & .... & .... \\ \widehat{\gamma}_{1n} & \widehat{\gamma}_{2n} & .... & \widehat{\gamma}_{nn} \end{pmatrix} \tag{18}$$

The random vector y and the parameter vector θ_0 satisfy the $\frac{n(n+1)}{2}$ population moment conditions:

$$E\left[f(y,\theta_0)\right] = \begin{pmatrix} \begin{array}{r} \widehat{\gamma}_{ii} - \sigma_i^2 - \Sigma_{j\neq i}^n \sigma_j^2 \omega_{ij}^2 \\ i = 1, 2..., n \end{array} \\ \begin{array}{r} \widehat{\gamma}_{ij} - \omega_{ij}\left(\sigma_i^2 + \sigma_j^2\right) - \sum_{\substack{l\neq i \\ l\neq j}}^n \sigma_l^2 \omega_{il}\omega_{jl} \\ i, j = 1, 2..., n; i < j \end{array} \end{pmatrix} = 0 \tag{19}$$

The model is exactly identified and can be solved according to the well-known Pearson's *Method of Moments*. There is a unique solution but the model is non-linear, which can hinder its resolution specially with increasing values of n.

The situation is more interesting if the SMA process of (16) has constant variance, so that matrix $\mathbf{\Sigma}$ can be written as $\sigma^2\mathbf{I}_n$. Introducing this restriction in (19) we obtain the following set of $\frac{n(n+1)}{2}$ population moment conditions:

$$g(\theta_0) = E\left[f(y, \theta_0)\right] = \left(\begin{array}{c} \widehat{\gamma}_{ii} - \sigma^2 \left(1 + \Sigma_{j \neq i}^{n} \omega_{ij}^2\right) \\ i = 1, 2 \ldots, n \\ \widehat{\gamma}_{ij} - \sigma^2 \left(2\omega_{ij} + \sum_{\substack{l \neq i \\ l \neq j}}^{n} \omega_{il}\omega_{jl}\right) \\ i, j = 1, 2 \ldots, n; i < j \end{array} \right) = 0 \quad (20)$$

The number of unknown parameters in $\theta_0 = (\sigma^2, \omega_{12}, ..., \omega_{1n}, \omega_{23} ..., \omega_{n-1,n})$ is $\frac{n(n-1)}{2} + 1$ so there are $(n-1)$ overidentifying restrictions. The problem can be solved using GMM estimators:

$$\underset{\theta}{Min} Q_T(\theta) = \frac{1}{T^2} \left\{ \Sigma_{t=1}^{T} f(y_t, \theta)' \right\} \mathbf{M}_T \left\{ \Sigma_{t=1}^{T} f(y_t, \theta) \right\} \quad (21)$$

being $\mathbf{M}_T$ a positive semi-definite matrix which converges to a positive definite matrix of constants For the problem to be well defined we need the following set of assumptions:

- The sample space is $\mathbf{Y} \in \mathbb{R}^n$ and the random vectors $\{y_t;\ -\infty < t < \infty\}$ are strictly stationary.
- The space parameter is Θ and θ_0 is an interior point of Θ. The $\frac{n(n+1)}{2}$ population moment conditions $E\left[f(y, \theta_0)\right] = 0$ are fulfilled in this point.
- Global identification in the sense that $E\left[f(y, \overline{\theta})\right] \neq 0$ where $\overline{\theta} \neq \theta_0$. The condition can be split into two clauses: (i) there is a unique correspondence between the covariance matrix and the 'scaled' weighting matrix $\rho\mathbf{W}$, which was proven in Battacharjee and Jensen-Butler (2013) and (ii) the moment conditions only hold for θ_0, which is difficult to prove because of the nonlinear relation of the weights and the variance term.
- Regularity conditions on the first derivatives of the vector of moment conditions. Specifically, $\partial f(y, \theta)/\partial\theta'$ exists, it is continuous in Θ, and $E\left[\partial f(y, \theta)/\partial\theta'\right]$ exists and is finite.

The model in (16) is nonlinear so the first order condition for minimizing (21) yields to a system of nonlinear equations; in general:

$$\frac{1}{T^2} \left\{ \Sigma_{t=1}^{T} \frac{\partial f(y_t, \theta)}{\partial\theta'} \right\}' \mathbf{M}_T \left\{ \Sigma_{t=1}^{T} f(y_t, \theta) \right\} = 0 \quad (22)$$

whose solution is the set of GMM estimation. Under the assumptions above and for the case of finite (small) n and increasing T, these estimators are consistent, so:

$$\widehat{\theta}_T \xrightarrow{p} \theta_0 \tag{23}$$

Moreover, if the sample moments verify a Central Limit Theorem, the GMM estimators converge asymptotically to a normal distribution:

$$T^{1/2}\left[\widehat{\theta}_T - \theta_0\right] \xrightarrow{d} N(0; \mathbf{RSR}') \tag{24}$$

where $\mathbf{S}$ is the limiting covariance matrix of the sample moments, $\lim_{T\to\infty} Var\left[T^{-1/2}\Sigma_{t=1}^{T} f(y_t, \theta\right] = \mathbf{S}$, and $\mathbf{R}$ is a $\left\{\frac{n(n-1)}{2}+1\right\} \times \frac{n(n+1)}{2}$ matrix such that: $\mathbf{R} = \left(\mathbf{G}_0'\mathbf{M}\mathbf{G}_0\right)^{-1}\mathbf{G}_0'\mathbf{M}$. $\mathbf{M}$ is the convergence matrix of $\mathbf{M}_T$ and $\mathbf{G}$ is the convergence matrix of matrix of first derivatives of the sample moments, $\mathbf{G}_T(\theta_0) = \frac{1}{T}\left[\Sigma_{t=1}^{T}\frac{\partial f(y_t,\theta_0)}{\partial\theta'}\right]$; this matrix is $\frac{n(n+1)}{2} \times \left\{\frac{n(n-1)}{2}+1\right\}$ whose details can be found in Appendix I.

Hansen (1982) proves that there is there is an optimal choice for $\mathbf{M}_T$, which corresponds to in the inverse of the covariance matrix of the vector of population moment conditions, $\mathbf{M}_T = \left\{Var\left[T^{-1/2}\Sigma_{t=1}^{T} f(y_t, \theta\right]\right\}^{-1} = \mathbf{S}^{-1}$. This matrix produces, on average, GMM estimators with minimum variance, in which case:

$$T^{1/2}\left[\widehat{\theta}_T - \theta_0\right] \xrightarrow{d} N(0; \left(\mathbf{G}_0'\mathbf{S}^{-1}\mathbf{G}_0\right)^{-1}) \tag{25}$$

The covariance matrix of the sample moments is unknown, $\mathbf{S} = E\left[f(y_t, \theta)f(y_t, \theta)'\right]$, and should be estimated beforehand. Assuming that the sample moments $\{f(y_t, \theta)\}$ is a serially uncorrelated sequence the sample analog is a logic decision:

$$\widehat{\mathbf{S}}_T = \frac{1}{T}\Sigma_{t=1}^{T} f(y_t, \widehat{\theta}_T) f(y_t, \widehat{\theta}_T) \tag{26}$$

This is a consistent estimate, as shown by White (1994). For the case of an autocorrelated and/or heteroskedastic sequence of sample moments, the HAC framework offers a convenient way to produce consistent estimates under relatively weak assumptions on the structure of the process (Newey and West, 1994). In sum, as usual in the GMM literature, we finish with the iterated GMM estimator, so that for the i-th step:

i If $i = 1$: use a sub-optimal weighting matrix such as $\mathbf{M}_T = \mathbf{I}_K$ being $K = \frac{n(n+1)}{2}$ the number of restrictions. The estimator thus obtained, $\widehat{\theta}_T^{(1)}$, is consistent and allows us to obtain a provisional estimate of the covariance matrix, $\widehat{\mathbf{S}}_T^{(1)}$.

ii Obtain the corresponding GMM estimates by iterating the procedure, so that in the i-th step $\mathbf{M}_T = \left\{\widehat{\mathbf{S}}_T^{(i-1)}\right\}^{-1}$. We would need also of convergence rules to stop the process, so that if $\left\|\widehat{\theta}_T^{(i)} - \widehat{\theta}_T^{(i-1)}\right\| < \tau_\theta$ the algorithm has converged and the GMM estimator is $\widehat{\theta}_T = \widehat{\theta}_T^{(i)}$. On the contrary, the algorithm goes to the $(i+1)$ iteration.

4 Testing for changes in W

The assumption of *structural stability* plays an important role in the GMM estimation. It is clear that if this hypothesis does not hold, the estimates are not consistent and the inference will be missleading. The present Section centers on this assumption. It is meaning.is clear in the sense that it implies that (19) holds for the whole sample; however, the alternative is more difficult to articulate because you have to specify how and when the model changes. There are multiple ways by which a model can have breaks (variance, parameters, functional form, etc.), although we focus on the case of **W** for which there are few proposals in the literature (for example, Angulo *et al*, 2017a and b). In relation to when, the GMM literature has focused mainly on the case of a discrete change at a single known point in the sample, the so-called *break point*. However, since the work of Andrews (1993), there is a growing literature which considers also the case where the location of the break point is unknown. We are going to explore both ways, in relation to **W**, but considering only one break because of estimation constraints.

4.1 The case of a known break point

As indicated before, the null hypothesis states that the assumption (19) holds throughout the sample. The alternative hypothesis indicates that there is a (single) known break point, in period T_b where $1 \leq T_b \leq T$. and the user knows its location. Hall (2005) introduces a interesting distinction between identifying and overidentifying, that is very useful for us. The question is that we have $q = \frac{n(n+1)}{2}$ moment conditions to estimate $K = \frac{n(n-1)}{2}+1$ parameters in the homoskedastic case of (16). K of the q conditions are used to identify the parameters and are imposed in the estimation algorithm. The null hypothesis, related to the set of identifying conditions, reads as:

$$H_0^{I,KB}(T_b): \left.\begin{array}{l} \mathbf{P}_1(\theta_0)\left\{\mathbf{S}_1(\theta_0)^{-1}E_1\left[f(y_t,\theta_0)\right]\right\}=0;\ 1\le t\le T_b \\ \mathbf{P}_2(\theta_0)\left\{\mathbf{S}_2(\theta_0)^{-1}E_2\left[f(y_t,\theta_0)\right]\right\}=0;\ T_b+1\le t\le T \end{array}\right\} \tag{27}$$

where $\mathbf{P}_j(\theta_0)=\mathbf{F}_j(\theta_0)\left[\mathbf{F}_j(\theta_0)'\mathbf{F}_j(\theta_0)\right]^{-1}\mathbf{F}_j(\theta_0)'$, $\mathbf{F}_j(\theta_0)=\mathbf{S}_j(\theta_0)^{-1/2}\mathbf{G}_j(\theta_0)$ being $\mathbf{G}_j(\theta_0)=E_j\left[\frac{\partial f(y_t,\theta_0)}{\partial\theta'}\right]$ and $\mathbf{S}_1(\theta_0)=\lim_{T\to\infty}Var_1\left[T_b^{-1/2}\Sigma_{t=1}^{T_b}f(y_t,\theta)\right]$[1]. The subindex denotes the corresponding subperiod. The identification conditions imply that the least squares projection of the standardized moments, $\left\{\mathbf{S}_j(\theta_0)^{-1}E_j\left[f(y_t,\theta_0)\right]\right\}$ a $(q\times 1)$ vector, onto the column space of $\mathbf{F}_j(\theta_0)$, a $(q\times K)$ matrix, is equal to zero or that $rank\left(\mathbf{P}_j(\theta_0)\right)=K$ In sum, the hypothesis of (27) states that the identification conditions remains the same in both subperiods, which extends to the parameters (weights in our case). Following Andrews (1993), a Wald, WA, Lagrange Multiplier, LM, or Likelihood Ratio-type, DJ, tests can be used to test this form of structural stability:[2][3]

$$WA(T_b)=T\left[\widehat{\theta}_{1,T}-\widehat{\theta}_{2,T}\right]'V_{WA}(T_b)^{-1}\left[\widehat{\theta}_{1,T}-\widehat{\theta}_{2,T}\right] \tag{28}$$

$$LM(T_b)=T\frac{T_b}{T-T_b}\left\{\mathbf{g}_1(\widehat{\theta}_T)'\widehat{\mathbf{S}}_T^{-1}(\widehat{\theta}_{1,T})\mathbf{G}_T(\widehat{\theta}_T)\right\}'V_{LM}(T_b)^{-1}\left\{\mathbf{g}_1(\widehat{\theta}_T)'\widehat{\mathbf{S}}_T^{-1}(\widehat{\theta}_{1,T})\mathbf{G}_T(\widehat{\theta}_T)\right\} \tag{29}$$

$$DJ(T_b)=T\left\{\mathbf{J}_T(\widehat{\theta}_T;\widehat{\theta}_T;T_b)-\mathbf{J}_{1,2}(\widehat{\theta}_{1,T};\widehat{\theta}_{2,T};T_b)\right\} \tag{30}$$

where $\mathbf{J}$ means the Sargan (1958) test of overidentification restrictions[4].

The complementary overidentifying restriccions to (27) can be stated as:

$$H_0^{O,KB}(T_b): \left.\begin{array}{l} \mathbf{N}_1(\theta_0)\left\{\mathbf{S}_1(\theta_0)^{-1}E_1\left[f(y_t,\theta_0)\right]\right\}=0;\ 1\le t\le T_b \\ \mathbf{N}_2(\theta_0)\left\{\mathbf{S}_2(\theta_0)^{-1}E_2\left[f(y_t,\theta_0)\right]\right\}=0;\ T_b+1\le t\le T \end{array}\right\} \tag{31}$$

---

[1] Similarly $\mathbf{S}_2(\theta_0)=\lim_{T\to\infty}Var_2\left[(T-T_b)^{-1/2}\Sigma_{t=T_b+1}^{T}f(y_t,\theta)\right]$

[2] $V_{WA}(T_b)^{-1} = \frac{T_b}{T}\left\{\mathbf{G}_1(\widehat{\theta}_{1,T})'\widehat{\mathbf{S}}_1^{-1}(\widehat{\theta}_{1,T})\mathbf{G}_1(\widehat{\theta}_{1,T})\right\}^{-1} + \frac{T-T_b}{T}\left\{\mathbf{G}_2(\widehat{\theta}_{2,T})'\widehat{\mathbf{S}}_2^{-1}(\widehat{\theta}_{2,T})\mathbf{G}_2(\widehat{\theta}_{2,T})\right\}^{-1}$

[3] $V_{LM}(T_b)=\left\{\mathbf{G}_T(\widehat{\theta}_T)'\widehat{\mathbf{S}}_T^{-1}(\widehat{\theta}_T)\mathbf{G}_T(\widehat{\theta}_T)\right\}$

[4] $\mathbf{J}_T(\widehat{\theta}_T;\widehat{\theta}_T;T_b)=T\mathbf{g}_T(\widehat{\theta}_T)'\widehat{\mathbf{S}}_T^{-1}(\widehat{\theta}_T)\mathbf{g}_T(\widehat{\theta}_T)$

$\mathbf{J}_{1,2}(\widehat{\theta}_{1,T};\widehat{\theta}_{2,T};T_b)=T\left\{\frac{T_b}{T}\mathbf{g}_{1,T}(\widehat{\theta}_{1,T})'\widehat{\mathbf{S}}_{1,T}^{-1}(\widehat{\theta}_{1,T})\mathbf{g}_{1,T}(\widehat{\theta}_{1,T})+\frac{T-T_b}{T}\mathbf{g}_{2,T}(\widehat{\theta}_{2,T})'\widehat{\mathbf{S}}_{2,T}^{-1}(\widehat{\theta}_{2,T})\mathbf{g}_{2,T}(\widehat{\theta}_{2,T})\right\}$

where $\mathbf{N}_j(\theta_0) = \mathbf{I}_q - \mathbf{P}_j(\theta_0)$; $j = 1, 2$. These $q - K$ conditions have been ignored in the estimation but they have helped to improve estimation efficiency. In this case we are pointing to changes not only in the weights but that extend to other features of the model (type of spatial dependence, perhaps heteroskedasticity, functional form, etc.). Hall and Sen (1999) propose the following test statistic:

$$\begin{array}{l} \mathbf{O}_T(T_b) = \mathbf{O}_{1,T}(T_b) + \mathbf{O}_{2,T}(T_b) \\ \mathbf{O}_{1,T}(T_b) = T\left\{\frac{T_b}{T}\mathbf{g}_{1,T}(\widehat{\theta}_{1,T})'\widehat{\mathbf{S}}_T^{-1}(\widehat{\theta}_{1,T})\mathbf{g}_{1,T}(\widehat{\theta}_{1,T})\right\} \\ \mathbf{O}_{2,T}(T_b) = T\left\{\frac{T-T_b}{T}\mathbf{g}_{2,T}(\widehat{\theta}_{2,T})'\widehat{\mathbf{S}}_T^{-1}(\widehat{\theta}_{2,T})\mathbf{g}_{2,T}(\widehat{\theta}_{2,T})\right\} \end{array} \tag{32}$$

The three identification structural stability test have an asymptotically $\chi^2(K)$ distribution whereas the stability overidentification test of (32) follows asymptotically a $\chi^2(q-K)$. More important, both set of tests (identification and overidentification) are asymptotically independent. In sum, the null hypothesis of structural stability must combine both clauses of identification (parameters and weights in $\mathbf{W}$) and overidentification stability (structure of the model), therefore we can state, for the case of a known break point, that:

$$H_0^{KB}(T_b) : H_0^{I,KB}(T_b) \,\&\, H_0^{O,KB}(T_b) \tag{33}$$

4.2 The case of an unknown break point

In this case, we are interesting in testing if there are symptoms of instability in any point in the sample, but we have not any a priori. To begin with, let us introduce some notation that it is usual in the literature of structural breaks. First, we normalize the break point space to the range $(0, 1)$, by defining a new parameter π, $0 \leq \pi \leq 1$ such that $T_b = [\pi T]$, where $[-]$ means the integer part. Moreover, it is habitual to trim this range to an smaller interval to avoid unexpected results at both extremes of the interval for π; $\pi \in \Pi = [0, 15; 0.85]$ is a rather common decision. Continuing with the discussion in last subsection, we can write the null hypothesis for this case:

$$H_0^{UB}(\pi) : H_0^{I,UB}(\pi) \,\&\, H_0^{O,UB}(\pi);\ \pi \in \Pi = [0, 15; 0.85] \tag{34}$$

The meaning and interpretation of the two clauses of the null coincide with that of (33). The testing sequence is a natural extension of the fixed break point framework, but obtaining the corresponding statistics for each possible value of π in the selected range, Π. This produces a sequence of statistics indexed by π and inference can be based on such sequence. Andrews and Ploberger (1994) suggest three alternatives, among which the *Supremum* value of the sequence is the most popular (because, for example, the value of $\widehat{\pi}$ where the sequence attains its maximum, appears to be a consistent estimation of the break point fraction). For example, focusing on the identifying restrictions, $H_0^{I,UB}(\pi)$, and the Likelihood Ratio-type test, DJ, they show that, under general conditions, the optimal statistic is:

$$Sup\ DJ_T = \sup_{\pi \in \Pi} \{DJ(\pi)\} \tag{35}$$

The limiting distribution of the $Sup\ DJ_T$ statistic has been tabulated in Andrews and Ploberger (1994); the same applies for the Wald and the Lagrange Multiplier statistics as well as for the *Average* and *Exponential* versions of the Supremum statistic. The $\mathbf{O}_T$ statistic of (32) can be extended, along the same lines, for testing the stability of overidentifying conditions when the location of the break point is unknown. Hall and Sen (1999) suggest obtaining the sequence of $\mathbf{O}_T$ values, indexed by π, and use the corresponding *Supremum* function (other alternatives are the *Average* and the *Exponential* functions), whose limiting distribution they tabulate

$$Sup\ \mathbf{O}_T = \sup_{\pi \in \Pi} \{\mathbf{O}_T(\pi)\} \tag{36}$$

Finally, the strategy proposed by Hall and Sen (1999) to detect the source of instability in the model is as follows:

1. If the unknown break point tests fail to reject the null hypothesis, there is evidence that the model is stable, both in parameters (weight matrix) and structure of dependence.

2. If the unknown identifying tests reject the null hypothesis and the unknown overidentifying tests are significant, there is evidence of parameter variation: the weight matrix is instable but the structure of dependence remains the same.

3. All other cases should be interpreted in the sense that there is instability that involves more than the parameters alone.

5 Description of the Monte Carlo

This part of the paper is devoted to the Monte Carlo conducted in order to calibrate the performance of the GMM procedure to estimate an unknown composite weighting matrix **W** and to test for their stabilitity over the sampling range. The procedures described in previous Sections are very demanding in terms of computing so our focus is on small sample sizes; specifically all of our experiments involve a large T and a very small n. Our purpose is to extend the experimental framework in the future.

We are going to simulate simple models, like the SMA in (16) or the correspondig SAR, in a panel setting; for the moment, we do not introduce unobserved terms in the panel and the analysis is confined to the homoskedastic case. For the SMA case:

$$\begin{gathered} y_t = \varepsilon_t + \rho \mathbf{S} \varepsilon_t = \varepsilon_t + \mathbf{W} \varepsilon_t ;\ \varepsilon_t \sim N\left(0, \boldsymbol{\Sigma}\right) \\ t = 1, 2, ..., T \end{gathered} \tag{37}$$

y_t and ε_t are $(n \times 1)$ vectors and $\mathbf{S}$ a $(n \times n)$ matrix to be define below;ρ is the unidentified parameter of spatial dependence. The error terms are obtained from a normal distribution: $\varepsilon_{it} \sim i.i.d.N\left(0; \sigma_\varepsilon^2\right)$, so that $\boldsymbol{\Sigma} = \sigma_\varepsilon^2 I_n$ with $\sigma_\varepsilon^2 = 1$.

As said, the weighting matrix, $\mathbf{S}$, is unknown and we are trying to estimate it from the data using the GMM procedures described before.We only assume that the weights in $\mathbf{S}$ are symmetrical, non-negative (this assumption is not required and it is made only to simplify) and equal to zero in the leading diagonal of the matrix, $s_{ii} = 0; i = 1, 2, ..., n$.

The main parameters of the experiment are the following:

- Only four different, small cross-sectional sample sizes, n, have been used $n \in \{4, 8, 12, 16\}$. The number of parameters that we have to recover from the GMM algorithm are, respectively, $K = \{7, 29, 67, 121\}$ so K is proportional to n^2 , $K = \frac{1}{2}O(n^2)$.

- The number of cross-sections in the panel, T, is intentionally large because of GMM requirements, $T \in \{500, 1000\}$.

- The values for the coefficient of spatial dependence, ρ, take only nonnegatives values, $\rho = \{0.2, 0.5, 0.8\}$.

- Finally, the weights in $\mathbf{S}$ are obtained from a uniform distribution, $s_{ij} \sim U(0, 1); i, j = 1, ..., n;\ j > i$.

- Each case has been replicated 100 times.

The second part of the experiment is devoted to testing for the hypothesis of structural stability in the model underlying the data. For the moment, we have preliminary results only for the case of a known break point (the unknown break point case is still under process). The break point is always located in the middle of the sample, so $T_b \in \{250, 500\}$ and the break may refer to three cases:

(i) Case 1. Change in the weighting matrix alone so that in the first period, $t \leq T_b$, intervene a set of weights different from that that act in the second period, $t > T_b$. Different means a differente draw from the Uniform distribution; the value of the coefficient of spatial dependence, ρ, has been keept fixed at 0.8.

(ii) Case 2. Change in the mechanism of spatial dependence from SMA in the first period to SAR in the second. The weigthing matrix and the coefficient of spatial dependence remain the same (0.8 the second).

(iii) Case 3. Change in both aspects, the matrix and the mechanism of spatial dependence, combining cases (i) and (ii).

5.1 Results of the Monte Carlo

Tables 1 and 2 summarize the main results of our Monte Carlo in relation to the recovery, using GMM estimates, of the unobserved weights. The tables show the Average Entropy Loss (EL) and the Average Frobenius Loss (FL), averaging the 100 experiments for each case. The EL and FL are defined as (Moscone *et al.*, 2017):

$$EL = tr\left(\mathbf{B}^{-1}\widehat{\mathbf{B}}\right) - \log(abs\left|\mathbf{B}^{-1}\widehat{\mathbf{B}}\right|) - n \qquad \text{(mc2)}$$

$$FL = \frac{\left\|\mathbf{B} - \widehat{\mathbf{B}}\right\|_F^2}{\|\mathbf{B}\|_F^2} \qquad (38)$$

where $\mathbf{B} = \mathbf{I}_n - \mathbf{W} = \mathbf{I}_n - \rho\mathbf{S}$ and $\|-\|_F^2$ means the Frobenius norm. The two indicators with take lower values the better the GMM estimation of the weighting matrix.

Table 1: Accuracy of the GMM estimates of the weights SMA case

				$T = 500$				
	$n = 4$		$n = 8$		$n = 12$		$n = 16$	
ρ	EL	FL	EL	FL	EL	FL	EL	FL
0.2	1.2411	0.3528	1.6220	0.3888	1.8447	0.4288	2.0754	0.4921
0.5	0.7651	0.2565	1.3107	0.2772	1.6448	0.3087	1.9660	0.3321
0.8	0.5548	0.1567	1.0689	0.1998	1.2148	0.2199	1.7655	0.2456
				$T = 1000$				
	$n = 4$		$n = 8$		$n = 12$		$n = 16$	
ρ	EL	FL	EL	FL	EL	FL	EL	FL
0.2	0.2287	0.1220	0.3988	0.1632	0.4399	0.2076	0.6543	0.2511
0.5	0.1642	0.0981	0.3290	0.1555	0.3587	0.1865	0.5446	0.2076
0.8	0.0987	0.0531	0.2456	0.1131	0.2875	0.1543	0.4124	0.1780

Overall the results are as expected. In first place, it is clear that the accuracy of the GMM algorithm improves strongly with T, so that doubling this parameter from 500 to 1000 implies a reduction in the accuracy indicators by more than half of their previous value. Lower values of T (smaller than 500) produce, in general, inaccurate estimates except in the case of a very low number of spatial units (4 to 8). Contrary, the increase of the cross-sectional size leads to a growing complexity for the GMM algorithm and worse results in terms of accuracy;the worsening in the accuracy indicators is a little less inelastic that the improvement observed in the case of an increase of T. The impact of the spatial coefficient is also positive in the sense that the accuracy of the estimates improves for higher values in this parameter. This result is somewhat unexpected because the indicators used, EL and FL, are free-scale measures of discrepancy; apparently the magnitude of the covariance terms (which increase with ρ) has a beneficial impact on the functioning of the GMM algorihm. Finally, the procedure work a little better in the case of SAR processes; this is also unexpected because the only difference between the two cases lies in the inversion of the covariance matrix, as indicated in (13).

Table 2: Accuracy of the GMM estimates. of the weights SAR case

	$T = 500$							
	$n = 4$		$n = 8$		$n = 12$		$n = 16$	
ρ	EL	FL	EL	FL	EL	FL	EL	FL
0.2	1.1261	0.3111	1.4864	0.3762	1.6487	0.4288	1.8901	0.4921
0.5	0.6321	0.2423	1.2975	0.2464	1.3486	0.3087	1.7265	0.3321
0.8	0.3897	0.1426	0.9453	0.2081	1.0404	0.2199	1.4586	0.2456
	$T = 1000$							
	$n = 4$		$n = 8$		$n = 12$		$n = 16$	
ρ	EL	FL	EL	FL	EL	FL	EL	FL
0.2	0.1196	0.1096	0.1334	0.1603	0.2654	0.2212	0.5880	0.2563
0.5	0.0654	0.0801	0.1187	0.1514	0.2097	0.1804	0.3502	0.2002
0.8	0.0425	0.0335	0.1456	0.1043	0.1775	0.1437	0.2201	0.1521

As said, the second part of the Monte Carlo deals with testing the assumption of structural stability in the spatial panel data model of (16). Our preliminary results refer to the case of a known break point located in the middle of sample, $T_b \in \{250, 500\}$,with three types of structural break (in the weights, in the form of the mechanism of spatial dependence and in both aspects). The discussion is solved using two statistics:

- the $DJ(T_b)$ Likelihood Ratio-type statistic of (30), asymptotically distributed as a $\chi^2(K)$ being $K = \frac{n(n-1)}{2} + 1$ to test the null hypothesis of (27) of stability in the identifying conditions (that is, estimated parameters).

- the $\mathbf{O}_T(T_b)$ statistic of (32), asymptotically distributed as a $\chi^2(q-K)$ being $K = \frac{n(n-1)}{2} + 1$ and $q = \frac{n(n+1)}{2}$ to test the null hypothesis of (31) of stability in the overidentifying conditions (that is, structure of the model).

According to the strategy suggested by Hall and Sen (1999), and described at the end of Section 4.2, we are going to accept the composite null hypothesis of structural stability of (33) if both statistics, $DJ(T_b)$ and $\mathbf{O}_T(T_b)$, are statistically not significant. A second possibility of interest emerges for the case of a $DJ(T_b)$ statistics statistically significant combined with a low $\mathbf{O}_T(T_b)$ statistic, not significant, then the evidence points to a problem of structural instability in the weights but not in the mechanisms of spatial dependence. All other combinations should be interpreted as revealing a problem of instability that involves more than just the weights.

Main results are summarized in Table 3, which shows the percentage of experiments with instability, simulated according to CASE 1, 2 or 3 that have been classified as Structural Stability (SS), Structural Instability in the Weights (SIW) or Structural Instability (SI). At the top of each panel appears the estimated size

Table 3: Structural stability test

					$T = 500$					
	SIZE	CASE 1			CASE 2			CASE 3		
n		SS	SIP	SI	SS	SIP	SI	SS	SIP	SI
4	0.09	0.03	0.82	0.15	0.04	0.06	0.90	0.02	0.00	0.98
8	0.12	0.01	0.74	0.25	0.01	0.07	0.92	0.01	0.01	0.98
12	0.15	0.00	0.65	0.35	0.00	0.09	0.91	0.02	0.03	0.95
					$T = 1000$					
	SIZE	CASE 1			CASE 2			CASE 3		
n		SS	SIP	SI	SS	SIP	SI	SS	SIP	SI
4	0.06	0.03	0.87	0.10	0.04	0.02	0.94	0.01	0.01	0.98
8	0.06	0.02	0.80	0.18	0.01	0.04	0.95	0.01	0.02	0.97
12	0.09	0.05	0.71	0.24	0.00	0.08	0.92	0.03	0.00	0.97

Overall, the results appear quite satisfactory for the case of a small number of cross section and a large value of T. The testing strategy is slightly oversized for the case of $T = 500$ but this excess tends to be corrected with large values of T. The Hall and Sen strategy works reasonable well when the stability problem affects only to the weights of W; then the percentage of correct identifications (Case 1 experiment classified as SIP) fluctuates around 80%. Let us note that the estimated probability that a Case 1 experiment is classified as SS (structural stability) remains quite low, below the 5% threshold. As expected, Cases 2 and 3 tend overwhelmingly to be classified as SI of structural instability.

6 Conclusions and future prospects

.

References

[1] Andrews, A.D. (1993): Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, **61**, 821-856.

[2] Andrews, A.D. and W. Ploberger (1994): Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative. *Econometrica*, **62**, 1383–1414.

[3] Angulo, A., P. Burridge and J. Mur (2017a): Testing for a structural break in the weight matrix of the Spatial Error or Spatial Lag model. *Spatial Economic Analysis*, **12**, 161-181.

[4] Angulo, A., P. Burridge and J. Mur (2017b): Testing for breaks in the weighting matrix. *Working Paper,* Economic Analysis Department, University of Zaragoza.

[5] Battacharjee, A. and C. Jensen-Butler (2013): Estimation of the Spatial Weights Matrix under Structural Constraints. *Regional Science and Urban Economics*, **43** 617-634.

[6] Corrado, L. and B. Fingleton (2012): Where is the Economics in Spatial Econometrics?. *Journal of Regional Science*, **52** 210–239.

[7] Hall, A. (2005): *Generalized Method of Moments*. Oxford: Oxford University Press.

[8] Hall, A. and A. Sen (1999): Structural Stability Testing in Models Estimated by Generalized Method of Moments. *Journal of Business and Economic Statistics*, **17** 335-348.

[9] Hansen, H. (1982): Large sample properties of Generalized Method of Moments estimators. Econometrica, **50** 1029-1054.

[10] Harris, R., J. Moffat and V. Kravtsova (2011): In search of 'W'. *Spatial Economic Analysis* **6**, 249-270

[11] Kelejian, H. and G. Piras (2014): Estimation of spatial models with endogenous weighting matrices, and an application to a demand

model for cigarettes. *Regional Science and Urban Economics*, **46** 140-149.

[12] Moscone, F., Tosetti, E. and Vinciotti, V. (2017): Sparse estimation of huge networks with a block-wise structure. *The Econometrics Journal*. doi:10.1111/ectj.12078

[13] Newey, W. and K. West (1994): Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, **61** 631-653.

[14] Pesaran, H. (2004): General diagnostic tests for cross section dependence in panels. University of Cambridge, Faculty of Economics, Cambridge Working Papers in Economics No. 0435.

[15] Sargan, J. (1958): The estimation of economic relationships using instrumental variables. *Econometrica*, **26**, 393-415.

[16] Snijders, T., C. Steglich and M.Schweinberger (2007): Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud and A. Satorra (eds) *Longitudinal models in the behavioral and related sciences*, pp. 41-71. London: Lawrence Erlbaum.

[17] White, H. (1994): *Estimation, Inference and Specification Analysis.* New York: Cambridge University Press.

Appendix I: Matrix of first derivatives of the vector of sample moments.

$$
\frac{\partial f(\nu,\sigma^2)}{\partial\theta'}=\left(\begin{array}{cccccccccccccccccccc}
-(1+\Sigma\omega_{1j}^2) & -2\sigma^2\omega_{12} & -2\sigma^2\omega_{13} & \cdots & -2\sigma^2\omega_{1n} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
-(1+\Sigma\omega_{2j}^2) & -2\sigma^2\omega_{12} & 0 & \cdots & 0 & -2\sigma^2\omega_{23} & -2\sigma^2\omega_{24} & \cdots & -2\sigma^2\omega_{2n} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
-(1+\Sigma\omega_{nj}^2) & 0 & 0 & \cdots & -2\sigma^2\omega_{1n} & 0 & 0 & \cdots & -2\sigma^2\omega_{2n} & 0 & 0 & \cdots & -2\sigma^2\omega_{3n} & \cdots & 0 & 0 & -2\sigma^2\omega_{n-3n} & 0 & -2\sigma^2\omega_{n-2n} & -2\sigma^2\omega_{n-1n} \\
-(2\omega_{12}+\Sigma\omega_{1k}\omega_{2k}) & -2\sigma^2 & -\sigma^2\omega_{23} & \cdots & -\sigma^2\omega_{2n} & -\sigma^2\omega_{13} & -\sigma^2\omega_{14} & \cdots & -\sigma^2\omega_{1n} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
-(2\omega_{13}+\Sigma\omega_{1k}\omega_{3k}) & -\sigma^2\omega_{23} & -2\sigma^2 & \cdots & -\sigma^2\omega_{3n} & -\sigma^2\omega_{12} & 0 & \cdots & 0 & -\sigma^2\omega_{14} & -\sigma^2\omega_{15} & \cdots & -\sigma^2\omega_{1n} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
-(2\omega_{1n}+\Sigma\omega_{1k}\omega_{nk}) & -\sigma^2\omega_{2n} & -\sigma^2\omega_{3n} & \cdots & -2\sigma^2 & 0 & 0 & \cdots & -\sigma^2\omega_{12} & 0 & 0 & \cdots & -\sigma^2\omega_{13} & \cdots & 0 & 0 & -\sigma^2\omega_{1n-3} & 0 & -\sigma^2\omega_{1n-2} & -\sigma^2\omega_{1n-1} \\
-(2\omega_{23}+\Sigma\omega_{2k}\omega_{3k}) & -\sigma^2\omega_{31} & -\sigma^2\omega_{32} & \cdots & 0 & -2\sigma^2 & -\sigma^2\omega_{34} & \cdots & -\sigma^2\omega_{3n} & -\sigma^2\omega_{24} & -\sigma^2\omega_{25} & \cdots & -\sigma^2\omega_{2n} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
-(2\omega_{24}+\Sigma\omega_{2k}\omega_{4k}) & -\sigma^2\omega_{41} & 0 & \cdots & 0 & -\sigma^2\omega_{34} & -2\sigma^2 & \cdots & -\sigma^2\omega_{4n} & -\sigma^2\omega_{23} & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
-(2\omega_{2n}+\Sigma\omega_{2k}\omega_{nk}) & -\sigma^2\omega_{n1} & 0 & \cdots & -\sigma^2\omega_{21} & -\sigma^2\omega_{3n} & -\sigma^2\omega_{4n} & \cdots & -2\sigma^2 & 0 & 0 & \cdots & -\sigma^2\omega_{23} & \cdots & 0 & 0 & -\sigma^2\omega_{2n-3} & 0 & -\sigma^2\omega_{2n-2} & -\sigma^2\omega_{2n-1} \\
-(2\omega_{34}+\Sigma\omega_{3k}\omega_{4k}) & 0 & -\sigma^2\omega_{14} & \cdots & 0 & -\sigma^2\omega_{42} & -\sigma^2\omega_{23} & \cdots & 0 & -2\sigma^2 & -\sigma^2\omega_{45} & \cdots & -\sigma^2\omega_{4n} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
-(2\omega_{35}+\Sigma\omega_{3k}\omega_{5k}) & 0 & -\sigma^2\omega_{15} & \cdots & 0 & -\sigma^2\omega_{52} & 0 & \cdots & 0 & -\sigma^2\omega_{45} & -2\sigma^2 & \cdots & -\sigma^2\omega_{5n} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
-(2\omega_{3n}+\Sigma\omega_{3k}\omega_{nk}) & 0 & -\sigma^2\omega_{1n} & \cdots & -\sigma^2\omega_{3n} & -\sigma^2\omega_{n2} & 0 & \cdots & -\sigma^2\omega_{n4} & -\sigma^2\omega_{n5} & -\sigma^2\omega_{4n} & \cdots & -2\sigma^2 & \cdots & 0 & 0 & -\sigma^2\omega_{3n-3} & 0 & -\sigma^2\omega_{3n-2} & -\sigma^2\omega_{3n-1} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
-(2\omega_{n-3n-2}+\Sigma\omega_{n-3k}\omega_{n-2k}) & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & -2\sigma^2 & -\sigma^2\omega_{n-2n-1} & -\sigma^2\omega_{n-2n} & -\sigma^2\omega_{n-3n-1} & -\sigma^2\omega_{n-3n} & 0 \\
-(2\omega_{n-3n-1}+\Sigma\omega_{n-3k}\omega_{n-1k}) & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & -\sigma^2\omega_{n-2n-1} & -2\sigma^2 & -\sigma^2\omega_{n-1n} & -\sigma^2\omega_{n-3n-2} & 0 & -\sigma^2\omega_{n-3n} \\
-(2\omega_{n-3n}+\Sigma\omega_{n-3k}\omega_{nk}) & 0 & 0 & \cdots & -\sigma^2\omega_{n-31} & 0 & 0 & \cdots & -\sigma^2\omega_{n-32} & 0 & 0 & \cdots & -\sigma^2\omega_{n-33} & \cdots & -\sigma^2\omega_{n-2n} & -\sigma^2\omega_{n-1n} & -2\sigma^2 & 0 & -\sigma^2\omega_{n-3n-2} & -\sigma^2\omega_{n-3n-1} \\
-(2\omega_{n-2n-1}+\Sigma\omega_{n-2k}\omega_{n-1k}) & 0 & 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & -\sigma^2\omega_{n-3n-2} & 0 & -2\sigma^2 & -\sigma^2\omega_{n-1n} & -\sigma^2\omega_{n-2n} \\
-(2\omega_{n-2n}+\Sigma\omega_{n-2k}\omega_{nk}) & 0 & 0 & \cdots & -\sigma^2\omega_{n-21} & 0 & 0 & \cdots & -\sigma^2\omega_{n-22} & 0 & 0 & \cdots & -\sigma^2\omega_{n-23} & \cdots & -\sigma^2\omega_{n-3n} & 0 & -\sigma^2\omega_{n-2n} & -\sigma^2\omega_{n-1n} & -2\sigma^2 & -\sigma^2\omega_{n-2n-1} \\
-(2\omega_{n-1n}+\Sigma\omega_{n-1k}\omega_{nk}) & 0 & 0 & \cdots & -\sigma^2\omega_{n-11} & 0 & 0 & \cdots & -\sigma^2\omega_{n-12} & 0 & 0 & \cdots & -\sigma^2\omega_{n-13} & \cdots & 0 & -\sigma^2\omega_{n-3n} & -\sigma^2\omega_{n-3n-1} & -\sigma^2\omega_{n-2n} & -\sigma^2\omega_{n-2n-1} & -2\sigma^2
\end{array}\right)
$$